

Weekly Report

20171121-20171126

Han Dongming Nov. 26 2017

看了几篇query pricing的文章，发现他们都
开始上课了之后比较忙

Paper

Z-Glyph: Visualizing outliers in multivariate data

Voila: Visual Anomaly Detection and Monitoring with Streaming Spatiotemporal Data

Z-Glyph: Visualizing outliers in multivariate data

摘要

multivariate data

挑战

1. defining “normal” (and “anomalous”) behavior 定义异常，正常的行为
2. 高质量的标记数据难以获得

贡献

1. Extending the existing design. (Z-Star Glyph的拓展)
利用了
 - human perception features
 - visual metaphor 隐喻
 - statistical characterization
2. Extensive controlled experiment.
3. Case studies on real datasets.

相关工作

1. outlier detection
2. glyph-based visualization
3. similar visual designs

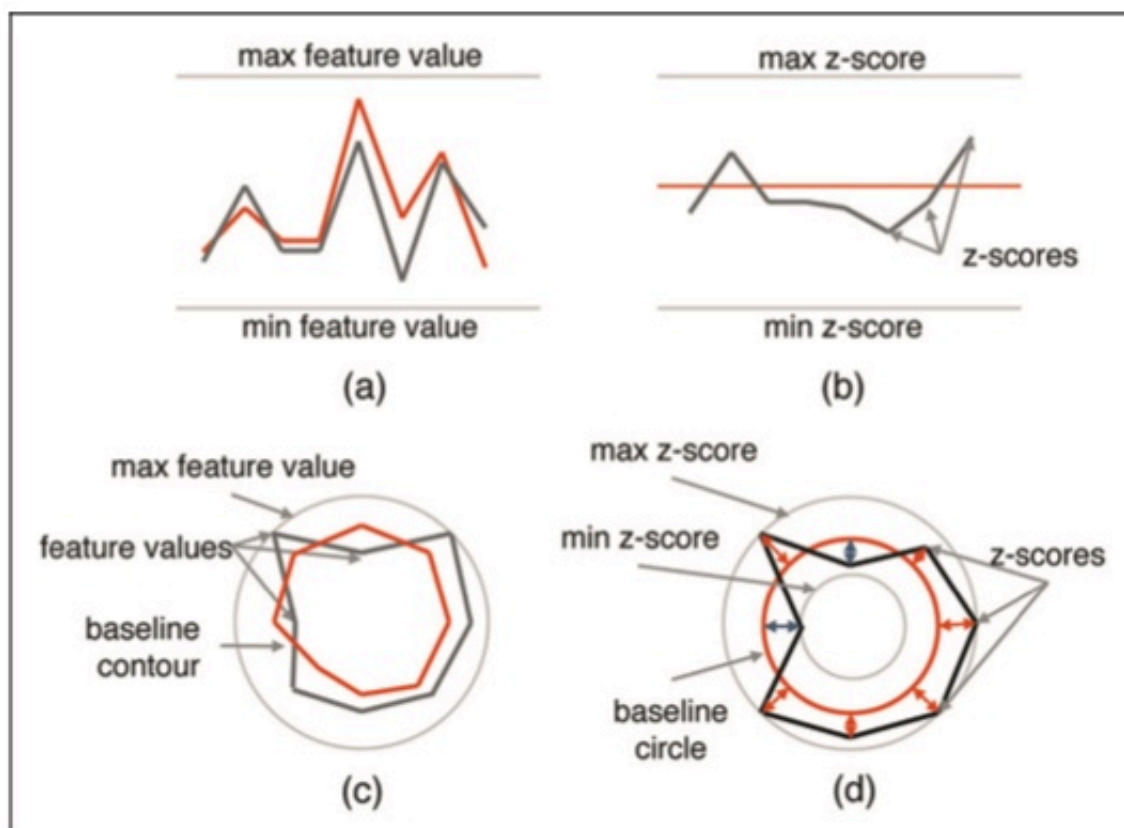
评估

1. Case studies
2. experts interview

设计

问题： how to represent outlier information that can be easily perceived and recognized by human

- Choosing optimal visual channels.
- Utilizing visual metaphor
- Incorporating statistical distribution concept.



讨论

Z-Glyph 更适合特征符合正态分布的数据

结论

the first set of glyphs that were designed for revealing outliers in a multivariate dataset.

Voila: Visual Anomaly Detection and Monitoring with Streaming Spatiotemporal Data

发表

VIS2017 VAST

Nan Cao, Chaoguang Lin, Qiuhan Yu-Ru Lin, Xian Teng, Xidao Wen

摘要

- **数据**: spatiotemporal data 时空数据的异常检测
- **现有的解决方案**: 限制在大型，动态，混合数据中识别异常
- **需求**: online monitoring and interactivity 在线检测和交互

挑战 与 需求

1. 适应性Adaptivity-----在线检测与分析Online monitoring and analysis.
2. 可解释性Interpretability----Multifaceted pattern discovery and anomaly filtering
3. 交互性Interactivity-----Human in the analysis loop

贡献

1. 系统，满足三个挑战
2. 算法，基于张量的异常检测
3. 可视化交互 Visualization and Interaction, 基于张量的异常检测算法，适用于动态输入数据以及展示pattern的spatiotemporal context，同时结合Bayesian approach去分析

相关工作

1. Anomaly Detection Algorithms
2. Visual Anomaly Detection
3. Visualizing the Spatiotemporal Data

系统

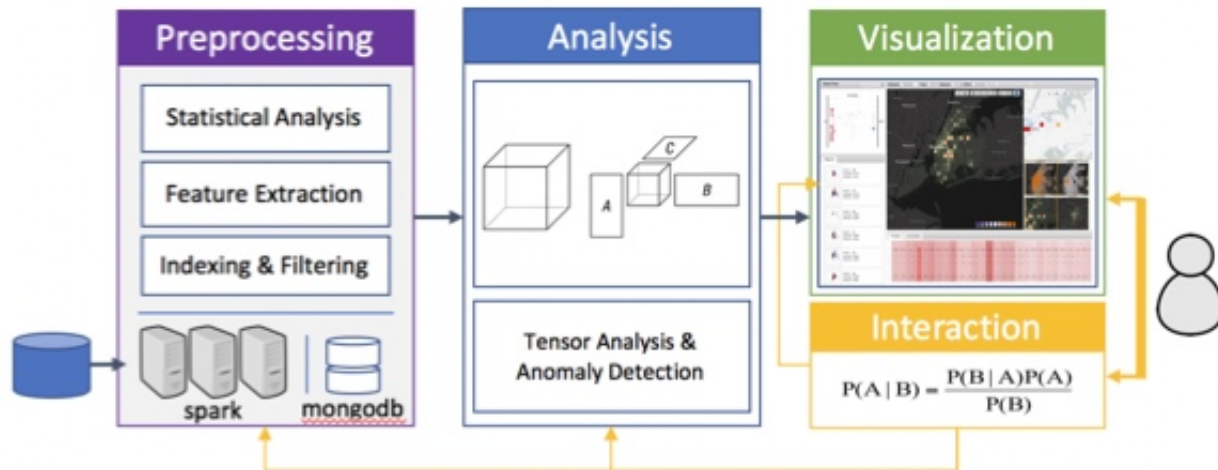


Fig. 2. The system architecture.

1. the data preprocessing module,
2. the analysis module,
3. the visualization module,
4. the interaction module

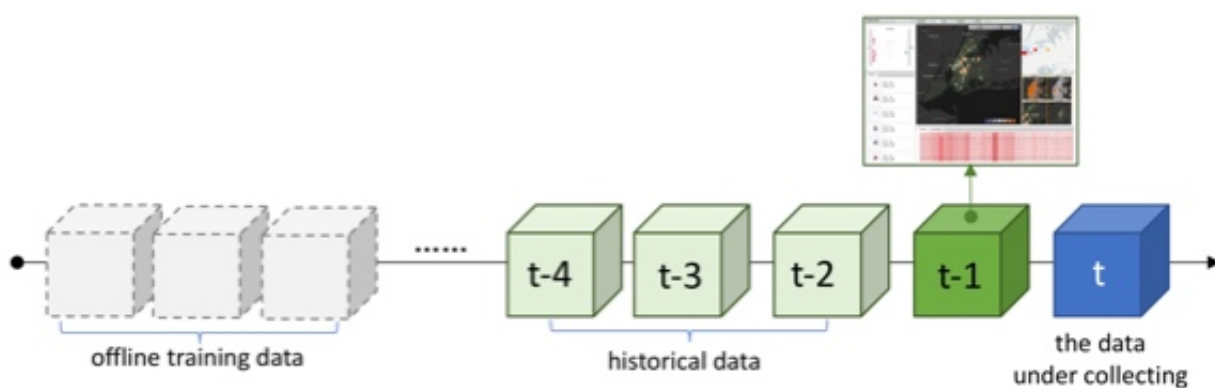


Fig. 3. The system pipeline for data processing and analysis.

异常检测算法

1. local outlier factor (LOF) 这篇中用LOF更好
2. One-Class SVM

数据集

New York City taxi-trip

various sources (e.g. taxi trips, traffic sensors, etc.)

可视化方法

Anomaly glyph.

small multiple charts

评估

Baseline methods and evaluation metrics

1. Case Study
2. Domain Expert Interview

算法

1. LOF
2. One-Class SVM

metrics

1. precision
2. recall
3. ROC

未来工作

1. providing tacit tutorials to guide the novice users
2. offering visual clues about low-precision instances in the anomaly ranking list
3. supporting fact search and checking
4. adaptively determining the temporal granularity,
5. developing new algorithms with forecasting and prediction capability.